

Research Article

An Integrated Support to Collaborative Semantic Annotation

**Annamaria Goy, Diego Magro, Giovanna Petrone, Claudia Picardi,
Marco Rovera, and Marino Segnan**

Dipartimento di Informatica, Università di Torino, C. Svizzera 185, 10149 Torino, Italy

Correspondence should be addressed to Annamaria Goy; annamaria.goy@unito.it

Received 24 September 2016; Revised 14 December 2016; Accepted 30 January 2017; Published 21 February 2017

Academic Editor: Thomas Mandl

Copyright © 2017 Annamaria Goy et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Everybody experiences every day the need to manage a huge amount of heterogeneous shared resources, causing information overload and fragmentation problems. Collaborative annotation tools are the most common way to address these issues, but collaboratively tagging resources is usually perceived as a boring and time consuming activity and a possible source of conflicts. To face this challenge, collaborative systems should effectively support users in the resource annotation activity and in the definition of a shared view. The main contribution of this paper is the presentation and the evaluation of a set of mechanisms (personal annotations over shared resources and tag suggestions) that provide users with the mentioned support. The goal of the evaluation was to (1) assess the improvement with respect to the situation without support; (2) evaluate the satisfaction of the users, with respect to both the final choice of annotations and possible conflicts; (3) evaluate the usefulness of the support mechanisms in terms of actual usage and user perception. The experiment consisted in a simulated collaborative work scenario, where small groups of users annotated a few resources and then answered a questionnaire. The evaluation results demonstrate that the proposed support mechanisms can reduce both overload and possible disagreement.

1. Introduction

Everybody, even less technologically skilled people, experiences every day the need to manage a huge amount of heterogeneous digital resources (documents, web pages, images, posts, emails, etc.): almost every kind of activity and service, in fact, has evolved from being based on a collection of paper documents, phone calls, and physical interactions to being fully digital. Public Administration services, stores, entertainment events, business interactions, reservations and payments, personal communications, and so on rely on web-based applications, accessible from desktop computers, tablets, smartphone, or other devices.

This trend poses a great challenge to individual users, who are often victims of both *information overload* and *information fragmentation* [1–3]: overload, since too many digital items (services, applications, contents, and resources) are available; fragmentation, because digital items are typically stored in different places, handled by different applications, encoded in different data formats, and accessed through different accounts. As a consequence, users have to manage

an increasing number of storages and folders, usually with similar names and related contents but handled by different applications [4].

Besides the challenge for individual users, another issue has emerged from the synergy between the cloud paradigm and the Social Web: knowledge sharing and web-based user collaboration. In many cases, in fact, the problem is not simply managing huge amounts of heterogeneous resources but doing it collaboratively, within shared virtual spaces, like social networks, collaboration tools, and shared workspaces. Collaborating in small groups or participating in large communities, people usually have different needs, goals, and perspectives, strongly influencing the way they organize information.

As we will try to show in detail in Sections 2 and 3, the idea of annotating information items, with tags and comments, is probably the most used way to address these issues: tags and comments, in fact, are the basic mechanism exploited by almost all web-based tools to support collaborative work, and in particular collaborative organization of shared resources.

However, the annotation activity is usually considered a heavy and time-consuming task. As a consequence, users are not encouraged to do it, or fail to do it properly, with the result that organization, and thus retrieval, of shared resources becomes even harder than without additional annotations.

To face this challenge, collaborative systems should become smart, in order to be able to effectively support users both in the individual activity of resource annotation and in the collaborative task of defining a shared, meaningful, and useful view over them.

The main contribution of this paper is the presentation and the evaluation of a set of mechanisms providing users with the mentioned support to the collaborative annotation of resources.

In the following, Section 2 will present the background related work; Section 3 will discuss the motivations and the overall goal of our approach; moreover, it will sketch the ontology exploited by our system (described in a previous work [5]). Section 4 will describe in detail the different mechanisms we propose to support collaborative semantic annotation of shared resources: a Tag Recommender (representing a novel contribution of this paper) and a Personal View Manager (whose implementation is described in a previous work [6]); Section 5 will present the user evaluation of the prototype implementing the proposed functionalities and represents the major contribution of this paper. Finally, Section 6 will conclude the paper.

2. Related Work

As already sketched in Section 1, the availability of an increasing number of heterogeneous digital resources and services led to an overload and fragmentation problem. The main aspects that must be taken into account by new interaction models that try to face these challenges emerge from the analysis of the limits of the *desktop metaphor*, conducted by Kaptelinin and Czerwinski [7], and can be summarized in (a) user *collaboration*; (b) organization of resources and people around an *activity*; (c) *uniform* management of *heterogeneous* objects and, in particular, *uniform annotations*. These aspects are partially supported by a number of approaches, with different emphasis on different features.

On a slightly different track, an interesting approach is represented by TellTable [8], a web-based framework for synchronous work in a virtual collaborative office; the work of Adler and colleagues is particularly interesting since it emphasizes the importance of providing users with the possibility of keeping a private section within the shared workspace (e.g., private notes).

As far as the organization around activities is concerned, several works can be mentioned. For example, Haystack [9] is built around the concept of customizable workspaces; the approaches based on the Activity-Based Computing paradigm [10] claim that thematic contexts should be built around user activities, and a similar principle is used in the Kimura system [11]; co-Activity Manager, built on top of Windows 7, is a more recent approach in the same direction, providing a cloud-based ubiquitous access [12].

A number of works explicitly face the issue of collaborative resource annotation: an interesting approach is represented by the work of Rau and colleagues [13], who present a hypertext annotation system in an e-learning perspective. Pearson and colleagues [14] describe a digital environment for collaborative reading, enabling note-taking in a way that mimics paper-based annotation which the users are used to. Jan and colleagues [15] present a web-based collaborative annotation system in which two types of filters are used in order to reduce the number of annotations on the basis of their quality; the authors show how annotation filtering can improve reading comprehension.

Not all of these approaches rely on semantic annotation and, even when they enable it, they lack an effective support for the annotation of resources. The integration of desktop-based user interfaces and semantic technologies is the goal of the *Semantic Desktop* initiative [16], mainly developed within the NEPOMUK project (<http://nepomuk.semanticdesktop.org/>). This approach aims at supporting the collaboration among knowledge workers, thanks to an open-source framework enabling the implementation of semantic desktops that integrate different applications and ontologies. A further development toward the integration of Web of Data sources can be found in [17].

The use of semantic knowledge to support users in the collaborative management of shared resources underlies many other approaches. For example, Koning and colleagues [18] describe the ontology used in the CineGrid Exchange network to support media data management. Uflacker and Zeier [19] present TCN (Team Collaboration Networks), a system for the analysis of patterns in collaboration activities within multidisciplinary design teams, based on an ontology that provides the vocabulary for describing interactions between people and/or information elements.

Strategies used to organize digital resources have been largely studied within the Personal Information Management field [20, 21]. The most popular approach to overcome the rigidity of traditional classification systems (e.g., folder hierarchies) is represented by tagging systems and folksonomies [22, 23]. Our approach is more structured than simple tagging systems, it does not aim at aggregating tags into a folksonomy, and it cannot be considered an actual crowdsourcing system, since it targets relatively small groups of people knowing each other and focusing on a specific collaborative activity. However, these systems provide interesting ideas and thus deserve some further discussion.

Folksonomies represent a user-centered view of the resource space, since they enable multifacets classifications by associating items with metadata representing different aspects (*facets*) of the shared resources [24]. Although very popular, also tagging systems and folksonomies have some limitations, as demonstrated by studies that compare the two models [25].

Interesting enhancements have been proposed by endowing tagging systems with semantic technologies [26, 27]. For example, FLOR [28] enriches folksonomies with online available semantic knowledge; GroupMe! [22] enables users to group content items related to a given topic; MOAT (*Meaning Of A Tag*) [29] allows the definition of shared

machine-readable meanings of tags, linked to Linked Open Data sets; SemKey [30] exploits Wikipedia and WordNet to provide semantic features to a collaborative tagging systems.

One of the main drawbacks of tagging systems is the fact that tags usually refer to different aspects of the tagged resource, often related to user goals; for example, tags can be used to define the resource type (“letter”), to describe the resource content (“seasonal diseases”), to provide an opinion (“good presentation”), and to link the resource to a task (“to read”) [31]. In order to account, in a systematic way, for this heterogeneity, a more structured semantic representation of resources is needed. A research field where such a structure is usually provided is Natural Language Processing (NLP). In tools used in the NLP community, in fact, annotations usually rely on a predefined schema that supports the association of tags with phrases within documents. The annotation schema and the vocabulary are usually provided by ontologies [32–34], which can define the metadata structure (e.g., Dublin Core: <http://www.dublincore.org/>) or provide domain-dependent vocabularies (e.g., the Getty Thesaurus of Geographic Names: <https://www.getty.edu/research/tools/vocabularies/tgn>, for the geographic domain). Interestingly, there are also collaborative versions of some NLP-oriented annotation tools [35, 36]. Useful surveys of (ontology-based) annotation frameworks can be found in [37, 38].

We conclude this section by mentioning a quite different concept of annotation, that is, the one implemented in tools explicitly aimed at supporting collaborative work by enabling users to add comments to digital documents (web pages, pdf documents, etc.) and share them; in these tools, the annotation is typically free text, with no semantic schema. The most popular example is Google Docs (<https://docs.google.com>); other examples are A.nnotate (<https://www.a.nnotate.com>), HyLighter (<http://www.hylighter.com/>), and Marquee (<https://www.marqueed.com>) for images.

3. Semantic Annotation in Collaborative Workspaces

As already stated, shared workspaces used by groups of people to collaboratively carry on different types of activities, belonging both to business and personal spheres, pose new challenges: the huge number of resources, formats, storages, applications, services, and accounts produces a significant overload problem, coupled with an information fragmentation issue [1–3]. The result is usually twofold: the user is lost in the information space and the management of resources turns into a largely inefficient activity.

The sharing opportunities supported by cloud infrastructures, social software, and tagging systems, although representing a great advantage, can—meanwhile—cause new problems.

In particular, tagging within collaborative resource management systems raises some major problems:

- (a) Tags are typically used to express very heterogeneous aspects of a resource (e.g., its type, its content, an opinion about it, or its use) [31].

- (b) Resource tagging is typically perceived as burdensome (boring and uninteresting) by users.
- (c) Different users, even though belonging to the same group (team or community), often have different perspectives over shared resources and thus tend to disagree about the proper tags that should describe resources.

In order to face issue (a), the resource management system should provide a more structured, machine-readable, semantic representation of resources, in line with ontology-driven approaches to resource annotation [38]. To this purpose, we defined an ontology, representing resources as *information objects* characterized by a set of properties: an information object can thus have an *encoding format* (e.g., *pdf*), an *author* (e.g., *Paris Municipality*), a *language* (e.g., *written French*), a *content*, represented by a main *topic* (e.g., *Paris*), and a number of *objects of discourse* (i.e., entities the resource “is about”; e.g., *Montmartre*, *Louvre*, and *Jardin du Luxembourg*). Moreover, an information object can have *parts*: for example, a web page typically has links to other pages or includes multimedia contents (images, videos, etc.).

The proposed ontology is based on existing models, namely, DOLCE [39], its extension OIO [40], and the Knowledge Module of O-CREAM-v2 [41]. It enables us to provide a structured and uniform representation of heterogeneous resources (documents, web pages, multimedia items, emails, etc.): for instance, thanks to the structure provided by the ontology, the tag describing the author of a document becomes the value of the *author* property, while tags describing the content become values for the *topic* and *object of discourse* properties.

A simplified example of the semantic representation of a web page (*wp*), according to the ontology, is shown in “*Simplified Example of a Semantic Representation of a Web Page*.” The predicates include a time parameter (*t*) that has been omitted in the OWL (<http://www.w3.org/TR/owl-features>) version of the ontology exploited in our prototype (see Section 4).

Simplified Example of a Semantic Representation of a Web Page

```
specifiedIn(wp, written_French, t)
hasEncodingSpecifiedIn(wp, pdf, t)
hasTopic(wp, Paris, t)
hasObjectOfDiscourse(wp, Montmartre, t)
hasObjectOfDiscourse(wp, Louvre, t)
hasObjectOfDiscourse(wp,
Jardin_du_Luxembourg, t)
hasAuthor(wp, Paris_Municipality, t)
hasURL(wp, http://www.myparis.org, t)
DOLCE:part(wp, image_JLux01, t)
DOLCE:part(wp, hyperlink_LouvreWebSite,
t)
```

The collaborative building the semantic representation of a resource can thus be seen as a “semantic annotation” activity, where annotations are represented by assertions

(e.g., *hasTopic(wp, Paris, t)*), based on the underlying ontology, in which the specific value (e.g., *Paris*) can be seen as a “tag.”

It is worth noting that the ontology provides the *structure* for the annotation by distinguishing different properties (topic, objects of discourse, author, format, etc.), but the values (and in particular the values of the *topic* and *object of discourse* properties) are free texts.

This solution enables users to organize resources on the basis of a highly structured information that provides a flexible access, based on the combination of different criteria (e.g., I can select all resources written in French by the Paris Municipality and talking about Montmartre).

The mentioned ontology has been implemented within our collaborative resource management system, SemT++ [42], and is described in detail in [5], where the reader can also find an evaluation of its benefits in performing resource selection.

An initial step toward the solution to issue (b) was provided by the identification of values for some of the mentioned properties, based on the automatic analysis of the resources. For example, by analyzing the HTML code of a web page, it is usually possible to understand its format (e.g., UTF-8/HTML) and sometimes its language (e.g., English); see [5] for details. However, the most challenging properties, with respect to user-burdening (issue (b)), but also interuser disagreement (issue (c)), are those referring to the *content* of the resource, that is, *topic* and *object of discourse*, which typically require to be manually provided by users.

In order to analyze the collaboration process and devise a strategy to support users in this task, and thus providing solutions to issues (b) and (c), we conducted a first preliminary user study—reported in [43]—where we asked users to collaboratively tag resources by focusing on the description of their content, with the goal of reaching a shared annotation, and without any specific support mechanism. Users were asked to write their annotations in a shared document (in particular, we used a Google Docs shared file) where everyone could freely edit others’ annotations, possibly deleting them. We monitored their process, without intervening. The aim of the study was to evaluate, among other things, the attitude people held toward the experience (users were asked to assess how interesting/engaging/easy/useful it was), as well as the degree of collaboration and/or disagreement perceived. Users tested different policies for deciding the final set of annotations (consensual, supervised by an external supervisor, or supervised by an internal owner of the resource). At the end of the collaborative process, users had to fill in a questionnaire where they evaluated the overall experience as well as their degree of satisfaction both with respect to the final result and with respect to the decision policy.

The study confirmed that users found the task not-so-engaging and only moderately interesting, in particular because of the initial difficulty in coming up with adequate tags and because of the difficulty in giving up one’s own view over the annotations in favor of someone else’s one. Also, the study showed that, while having the resource owner making the final decision is perceived as the most adequate policy,

it does not lead to a great level of satisfaction in the resulting annotations.

We built on the guidelines derived from such study by designing and implementing, within a proof-of-concept prototype, a set of mechanisms for effectively supporting users during the collaborative semantic annotation of shared resources (described in Section 4). We then performed an evaluation with users (reported in Section 5) in order to assess the degree of improvement with respect to the initial situation and to collect the users’ opinion about the support mechanisms.

4. An Integrated Support to Collaborative Semantic Annotation

In order to prototype and test our support mechanisms for collaborative annotation, we implemented them within our collaborative resource management environment, SemT++ [5, 42], in which shared workspaces are seen as *round tables*, devoted to specific activities (e.g., the organization of a holiday) and hosting resources of different types (documents, web pages, emails, images, etc.). Users “sitting” around a table (*table participants*) can collaboratively organize, retrieve, and use resources present on the table. These tasks are supported by the availability of semantic representations of resources themselves, based on the underlying ontology (see Section 3 and “Simplified Example of a Semantic Representation of a Web Page”), and in particular by the automatic identification of some resource properties, such as resource *parts* (basically, hyperlinks and multimedia objects), and all format-related properties. Moreover, the system tries to identify the language used and the authors, and, if it finds possible values (e.g., *written French* and *Paris Municipality*), it asks users for a confirmation. As we will see in Section 5, when describing the evaluation scenario, SemT++ uses the extracted parts to suggest new resources (e.g., linked pages) to table participants, who can select the most interesting ones and add them to the table.

However, as already mentioned, the most challenging properties, which require a significant user contribution, are those describing the content of the resource, that is, the main *topic* and the *objects of discourse* (entities the document is about). In order to support table participants in this annotation activity (i.e., in providing values for the *topic* and *object of discourse* properties), we designed and implemented two features, aimed at

- (i) reducing the overload in defining semantic tags (i.e., values for the mentioned properties)—that is, providing a partial solution to issue (b), as introduced in Section 3;
- (ii) handling possible disagreement among users—that is, providing a solution to issue (c), as introduced in Section 3.

These features are tag suggestion and personal views, and will be described in the following.

4.1. Suggestions. Our first hypothesis was that the most effective way to reduce the overload imposed by semantic annotation of resources is providing users with suggestions of content tags (i.e., values for *topic* and *object of discourse* properties). We thus built an integrated *Tag Recommender*, with the role of proposing semantic tags: when annotating a resource, users can accept a suggestion, by selecting the proposed value, or refuse it and add a new value.

The Tag Recommender has three components:

- (1) The *Resource Analyzer* extracts suggestions from the syntactic analysis of the resource itself. The current prototype only analyzes web pages and extracts the value of the *content* attribute of the *keywords* metatag of the HTML code.
- (2) The *Named Entity Extractor* provides suggestions on the basis of Named Entity Recognition (NER). In particular, in the current implementation, the Named Entity Extractor relies on Text Razor (<https://www.textrazor.com>), an NLP tool that includes a NER service offering a RESTful API for remote access. Text Razor analyzes the resource and extracts all Named Entities (e.g., *Paris*, *Tour Eiffel*, *Sorbonne*, etc.), each one associated with two attributes (among others): the estimated relevance of the entity within the document (a measure of “how on-topic or important that entity is to the document”) and a confidence score (“a measure of the engine’s confidence that the entity is a valid entity given the document context”); see Text Razor online documentation (<https://www.textrazor.com/docs/rest>). In the version of the prototype used for the evaluation presented in Section 5, we considered extracted entities with relevance > 0.5 (relevance range is $[0 \dots 1]$) and confidence > 5 (confidence range is $[0.5 \dots 10]$); such thresholds can be configured with different values, if needed.
- (3) The *Semantic Knowledge Manager* leverages the underlying ontology to infer candidate tags, by exploiting the reasoning module (currently, Facts++: <http://owl.cs.manchester.ac.uk/tools/fact/>). In particular, in the current version, it uses the *DOLCE*: *part*(x, z, t) property, thanks to a set of axioms enabling it to infer tags for a resource from tags used for its parts (e.g., hyperlinks included in it) and vice versa. For example, the following axiom states that if a resource x contains a part z , which has y as an object of discourse, then y is a candidate object of discourse for x .

$\text{InformationElement}(x) \wedge \text{DOLCE} : \text{part}(x, z, t)$

$\wedge \text{hasObjectOfDiscourse}(z, y, t)$ (1)

$\longrightarrow \text{hasCandidateObjectOfDiscourse}(x, y, t)$

As an example, the suggestions provided by the system for the *object of discourse* property belonging to the semantic representation of a web page about tourism in Ireland (<http://www.cliffsofmoher.ie/about-the-cliffs/obriens-tower>) are shown in the bottom part of Figure 1.



FIGURE 1: The property window of a table resource.

4.2. Personal Views. Our second hypothesis was that the availability of *personal views* over semantic annotations of shared resources can reduce possible conflicts deriving from disagreement about tags among collaborating users. Following this hypothesis, we designed and implemented Personal View Manager, that is, a mechanism enabling table participants to keep personal annotations over shared resources. A detailed description of the model and its implementation within SemT++ can be found in [6]; in the following we briefly summarize the most significant aspects, from the user viewpoint.

When a user decides to add a semantic annotation (e.g., an object of discourse, describing the content of a resource), they can decide to add it to the *shared view*, or to their *personal view*: in the first case, the tag will be visible to all table participants; in the second case, it will be only visible to themselves. Tags added to the personal view can be shared later on. Moreover, when looking at the tags added by somebody else in the shared view, the user can decide to mark them as “I like,” in order to include them also in their personal view; the main consequence is that if deleted from the shared view, such tags will remain available in their personal view.

Figure 1 shows the property window of a web page (<http://www.cliffsofmoher.ie/about-the-cliffs/obriens-tower>), in which the values for the *object of discourse* property are shown: bold face boxes represent shared annotations, the small heart is the marker meaning that the current user “likes” the annotation, and thin-face boxes identify “private” annotations that are visible only to the current user.

The personal view of an individual user over a shared resource thus includes “private” tags (thin-face boxes) and shared tags marked as “I like” (boldface boxes with small heart); in fact these last ones will be automatically turned into “private” tags if another participant deletes them from the shared view. The shared view of a table resource, from the perspective of each individual user, includes all shared tags, being marked as “I like” or not (i.e., all boldface boxes, with and without the small heart).

These functionalities are available for all property values that can be edited by users; values automatically set by the system (e.g., encoding formats) are by default assigned to the shared view and to all personal views (i.e., represented as bold face boxes with the small heart) and cannot be deleted. Finally, by right-clicking on a tag, table participants can see the author of an annotation.

5. User Evaluation

The goal of the user evaluation was the assessment of the support mechanisms described in Section 4 in relation to the problems of user-burdening (task perceived as boring, time-consuming, and often uninteresting)—issue (b) in Section 3—and interuser disagreement (conflict on annotations and no satisfaction in the results)—issue (c) in Section 3. More precisely our goal was to

- (1) evaluate the *attitude* users held *toward the experience* and assess the improvement with respect to the situation where users had no-support mechanisms [43];
- (2) evaluate the *satisfaction* of the users at the end of the process, both with respect to the final choice of annotations and with respect to possible conflicts in the group, again assessing the improvement with respect to the absence of support mechanisms;
- (3) evaluate the *usefulness of the support mechanisms* both in terms of actual usage and in terms of user perception.

Our hypothesis was that providing users with suggestions for the tags describing resource content (i.e., values of the *object of discourse* property) could make the task less burdening, while the availability of personal views could alleviate disagreements, allowing each participant to keep their own version of the annotations.

In order to test our hypothesis, we simulated a collaborative work situation where small groups of users had to annotate, by means of our tool, a few resources, focusing in particular on the *object of discourse* property, for which multiple values can be selected. In each group the resource “owner” (i.e., the user who added the resource to the table) had the final say on the annotations—as this had been perceived as the most adequate policy in the preliminary user study; see [43]. Each group had half an hour to an hour to perform the task, at the end of which each participant had to answer a questionnaire.

One of us was physically present with each test participant to record possible verbal comments and to help with the application user interface, which was novel to them. Moreover, their actions within the application were recorded.

5.1. Methodology. We recruited 15 participants among our colleagues, in order to have technology-acquainted people, familiar with collaborative tools, thus representing potential users of a collaborative environment like SemT++ (according to several authors (e.g., [44]) 15 is within the acceptable sample size range, albeit, admittedly, on the lower-bound side,

for qualitative evaluations, especially when the evaluation involves an in-depth analysis of a reasonably homogeneous group [45]; since our main goal was not a quantitative, statistically relevant evaluation, but rather a qualitative feedback from potential real users, we did not explicitly take into account the sample statistical representativeness).

We built 3 groups of 5 people each. All groups were asked to participate in the same scenario that was presented to the participants, in order to set the context, as follows: you are the organizers of a scientific workshop that will take place in Galway (Ireland). In order to carry on the organization activity, you set up a SemT++ table, on which you are currently discussing the destination of the social trip. You already informally talked about Cliffs of Moher and Aran Islands as possible destinations, but the decision is still open.

For each group, we identified a “leader” who had the task of selecting the resource to be added to the table and to be tagged during the test. In order to make the selection easier and less time-consuming, we preidentified a small group of web pages concerning interesting places in Ireland, not far from Galway. When the leader had selected the preferred resource and added it to the table, the system started the analysis and (among other things) extracted its *parts* (see Section 4), which—in the current version—correspond to images contained in the selected web page and hyperlinks referring to related resources. The extracted parts (images and related pages) were then suggested to the user, who could add the most interesting ones to the table (see Figure 2). In the test scenario, we asked each group leader to select 2 or 3 resources from this list.

At this point, the table was populated with 3 or 4 resources that could be tagged by table participants. Each user was asked to edit the values for the *object of discourse* property, which means adding/removing/sharing/liking tags describing the content of the resource, bearing in mind the overall goal (organizing the social trip for the Galway workshop). We also briefly showed them the features offered by the system, that is, the possibility of selecting values from the list of system suggestions, the possibility of maintaining a personal view, and the Like feature.

The tagging activity ended when the leader (i.e., the “owner” of the tagged resources) decided that they were satisfied with the tags, that is, with the semantic description of the resources.

After the tagging activity was completed, each participant was asked to fill in a questionnaire. The results of the user evaluation thus consist in the questionnaire answers, together with a log of the action performed by users while interacting with the application.

5.2. Results. The questionnaire consisted of four sections:

- (1) User profiling.
- (2) Overall quality of the experience.
- (3) User satisfaction.
- (4) Evaluation of support mechanisms.

In the *user profiling* section, we asked subjects to self-assess their familiarity with web tools and with collaborative tools,

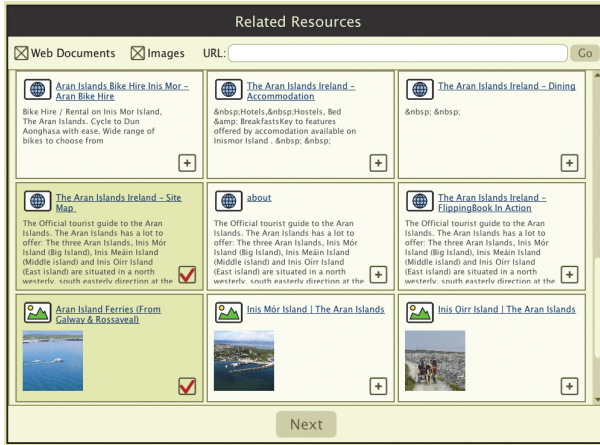


FIGURE 2: Selection of related resources suggested by the system.

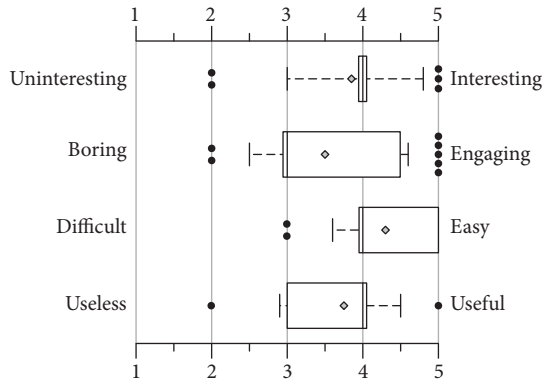


FIGURE 3: Assessment of the quality of the experience. Each boxplot shows the I, II, and III quartiles (box), the standard deviation (whiskers), outliers (black dots), and mean value (diamond).

on a 5-point scale. Most people declared a high or very high familiarity with web tools (4 people answered “4” and 11 people answered “5”; no one answered “1” to “3”); they were slightly less familiar with collaborative tools (9 people answered “3”; 5 people answered “4”; only one person answered “2”).

Concerning the *overall quality of the experience* (the second section in the questionnaire) we asked them to express on a 5-point scale four different quality indicators (the same used in the preliminary user study [43] in order to be able to fit them to the same grid): Uninteresting/Interesting, Boring/Engaging, Difficult/Easy, and Useless/Useful. Figure 3 shows a boxplot of the subjects’ answers.

The experience was considered *interesting* by most people (only 2 people gave a score lower than “4”) and, on the overall, more *engaging* than boring. The task was also perceived as very *easy* and reasonably *useful*.

Section 3 of the survey inquired on *user satisfaction* with respect to the final result and with respect to the decision policy (i.e., having the resource owner choose). Figure 4 shows the answers, again by means of boxplots. It can be seen that people were well satisfied with the chosen annotations

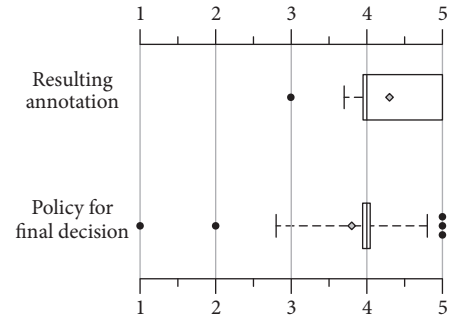


FIGURE 4: User satisfaction with respect to resulting annotations and decision policy.

(and also reasonably satisfied with the decision policy). We also explicitly asked if they had perceived conflicts within their group. Only 1 person out of 15 said she did.

Section 4 focused on the *evaluation of support mechanisms*. We asked users if they had used and/or found useful each of the following features:

- (i) The suggestions.
- (ii) The possibility of having personal tags invisible to others.
- (iii) The possibility of “liking” a tag.

Notice that the last two features both concur in creating the *personal view*; however, from the point of view of the interaction with our application, they could be accessed separately, and were, as a matter of fact, perceived as two different—albeit related—tools.

According to our test subjects, only one-third of them (5 people) explicitly interacted with the “private” annotations area (by adding, removing, or sharing a tag). Conversely, most of them said they used the Like feature (12 people). The action log confirms that only 5 people explicitly inserted a “private” tag, while another 4 actually interacted with the “private” area after a “private” tag had been generated due to the removal of a liked tag in the shared area. On the overall, 12.1% of the actions performed by our users concerned directly the personal annotations. However, 23.8% of the actions were either a “like” or “unlike,” and they were performed by 12 users out of 15, confirming the questionnaire answers.

Concerning system suggestions, we did not ask the subjects if they had “used” them because there were different ways of using the suggestions, for example, as a simple source of inspiration. Our action log, however, tells us that out of 106 tags added during the test, 19.8% were taken from those suggested by the system. Table 1 summarizes information on application usage.

The subjects’ answers on the usefulness of the different mechanisms are shown in Figure 5.

The last section of the survey focused more specifically on the suggestions provided by the system, asking subjects to assess their quality in terms of number, precision, and adherence to the resource topic, on a 3-point scale. Figure 6 shows the answers we obtained, represented as histograms.

TABLE 1: Summary of data on application usage logged during the evaluation.

(a)		
Total number of tag actions (add, remove, like, unlike, and share)	“Private” tag actions (add/remove from personal area, share a personal tag)	Like/unlike actions
206	25 (12.1%)	49 (23.8%)
(b)		
Total numbers of tags added	Tags added based on system suggestion	
106	21 (19,8%)	

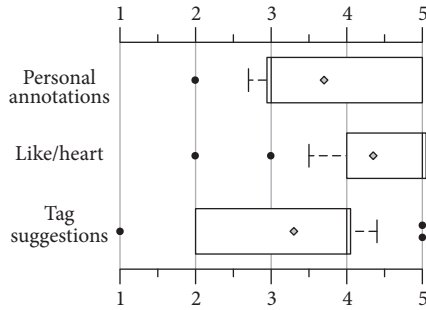


FIGURE 5: Users' assessment of the usefulness of each support mechanism.

Most people saw the suggestions as adequate in number (12) and in precision (10). No one thought the suggestions were too few, while 3 people thought they were too many. Among the 5 users not satisfied about the precision of suggestions, 2 found them too specific and 3 found them too general. While only 2 users found the suggestions off-topic, only 5 of them found them truly on-topic. Most users (8) saw them as moderately on-topic.

In the following section we discuss these results with respect to the initial goals of our investigation.

5.3. Analysis and Discussion. As discussed in the introduction of Section 5, our first goal was to “evaluate the attitude users held toward the experience and assess the improvement with respect to the situation where users had no-support mechanisms.”

In order to analyze this point, let us compare the users' answers on the quality of the experience in the two cases, namely, with or without support mechanisms. Figure 7 shows, for each quality indicator, the users' assessment in both situations.

It can be immediately noticed that there has been a significant improvement with respect to the Boring/Engaging indicator and a noticeable improvement also with respect to the Uninteresting/Interesting indicator. The other two indicators, Difficult/Easy and Useless/Useful, received a similar assessment in both questionnaires.

Although these answers do not specifically concern the support mechanisms, we can observe that including such

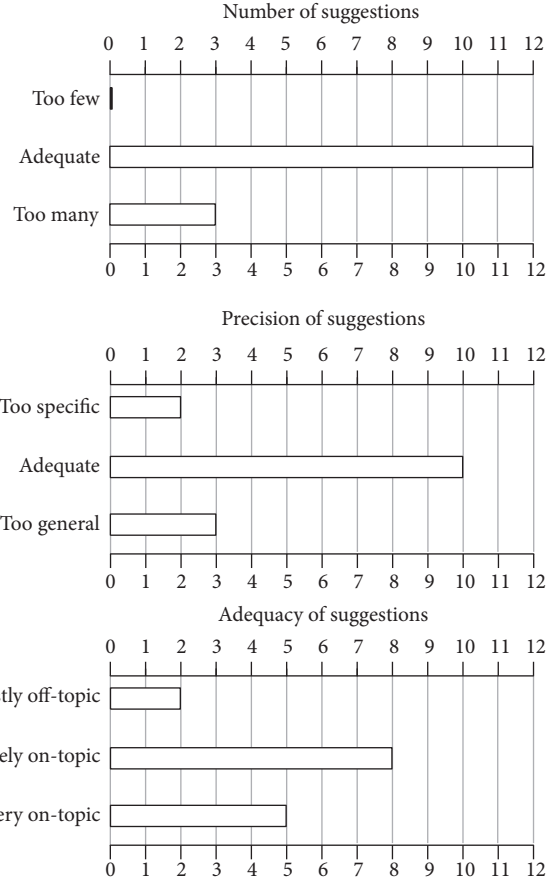


FIGURE 6: Quality of suggestions according to study subjects.

mechanisms in the prototype had a positive impact on the quality of the experience, in particular making it more engaging and interesting. It is worth noting that, in the user evaluation (II), the low outlier answers were given by the same people. In other words, while almost everyone found the experience at least not too difficult (no scores below 3 in the Difficult/Easy indicator), there were 2 people who were generally unsatisfied with the experience. Their free-text remarks lead us to think that they did not feel motivated or did not “see the practical utility” of the task they had been asked to perform. We can argue that the usefulness (or uselessness) of a task is somewhat subjective and does not depend on the tool used to perform it. The fact that the task was also seen by these people as somewhat boring and not so interesting can be seen as a consequence of this lack of motivation.

On the other hand, the fact that our application did not make things easier (referring to the Difficult/Easy indicator) for the participants can be ascribed to the intrinsic simplicity of the task itself. The task was indeed perceived as easy even in the preliminary study, where users had no specific tool to aid them.

Our second goal was to “evaluate the satisfaction of the users at the end of the process, both with respect to the final choice of annotations and with respect to possible conflicts

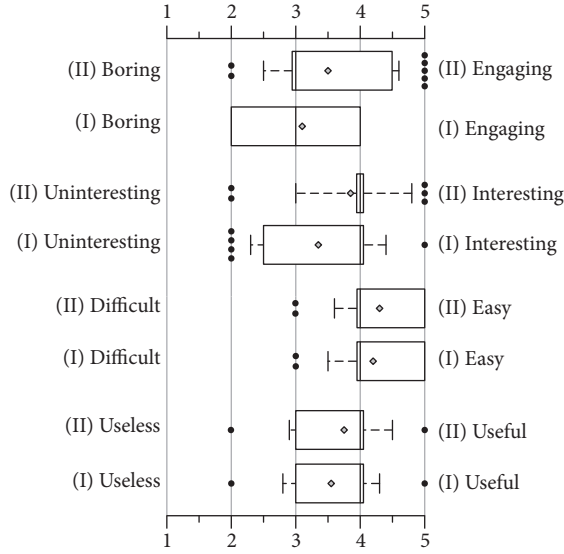


FIGURE 7: Comparison between questionnaire answers on experience assessment *with* (II) and *without* support mechanisms (I).

in the group, again assessing the improvement with respect to the absence of support mechanisms.”

In the evaluation presented in this paper we adopted an *authored* policy, which means that the resource owner (the person who initially chose it) had the final say on the annotations. In the preliminary study without support mechanisms we had in fact experimented with three different policies (authored, *consensual*, a decision was considered definitive only when everyone agreed upon it, and *supervised*, where an external supervisor, not participating in the annotation, had the final say).

In that case, the *authored* policy had been deemed the most adequate, but not the most satisfactory in terms of final annotation. Figure 8 compares the degree of satisfaction obtained in the user evaluation *with support mechanisms* to the score obtained by the three different policies in the case *without support mechanisms*.

The degree of satisfaction is similar (indeed, slightly higher) to the one obtained by the *consensual* policy without support mechanisms, definitely improving on the previous score for the *authored* policy. From the free-text comments of the subjects, we understand that the possibility of saving one’s own work in the personal view, rather than losing it if the resource owner decides to remove someone else’s tags, definitely contributes to this improvement in satisfaction.

Our third and last goal was to “evaluate the usefulness of the support mechanisms both in terms of actual usage and in terms of user perception.” The support mechanisms under investigation were the personal view and tag suggestions.

The personal view could be accessed by our users in two ways: by explicitly adding or removing private tags or by “liking” a shared tag, which had the effect of inserting it in the personal view too. Questionnaire answers reported in Figure 5, together with usage data in Table 1, show that users found the Like feature more useful and actually preferred to use it, rather than directly adding tags in the personal view.

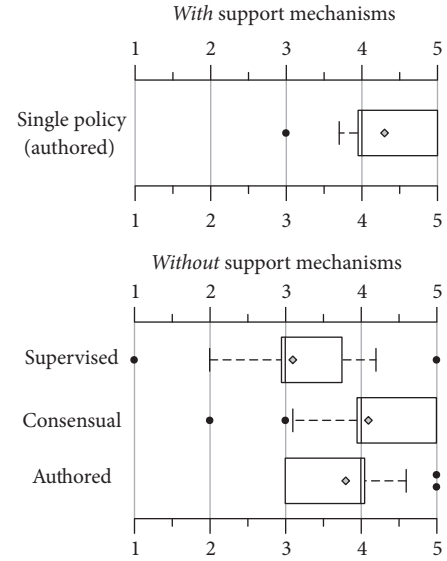


FIGURE 8: Comparison of satisfaction with resulting annotations between the evaluations *with* and *without* support mechanisms.

The free-text comments generally express the idea that the availability of a “private” annotation area is mostly considered useful for “saving” the preferred tags when they get removed by some other user. Interacting directly with it is, on the other hand, perceived by one-third of the users as redundant and unnecessary. Many participants pointed out that they saw “private” annotations as a sort of memo or back-up, and they should not be given in the user interface the same relevance as the shared ones (presently, the only difference is that shared tags are shown with a thick border, private ones with a thin one). A few users asked what was the point of adding personal annotations when the purpose was collaboration—they understood the idea of preserving removed tags, but beyond that they saw no use for the feature.

When asked explicitly about the Like/Unlike feature, most users noted that it allowed them to feel relaxed about the intervention of the other participants and in particular about the final choice of the resource owner. More than two-thirds of the participants (11 out of 15) said that they would have preferred the “likes” to be public (while no one was interested in others seeing their “private” tags), as a way of capturing the trend of the group. Some observed that, even in absence of an explicit voting mechanism, being able to see other people’s “likes” could help the resource owner to make the final choice, and in general help the group to understand what other people were at with their tags.

Concerning tag suggestions, we can see from Table 1 that about one-fifth of the tags were added following a hint from the system. The usefulness of the suggestion feature is however perceived as lower than the others. Some users (6 people) found the idea of receiving suggestions interesting but did not find that the suggestions themselves were adequate either in number, level of detail, or topic; 3 people reported that the problem was that suggestions focused on the content of the resource, while their tagging was goal-oriented; in other

words, they were interested in expressing the relationship between the content of the document and the purpose of their joint work. This type of tags could be more easily derived—in a real situation, where the table contained more resources—from tags associated with other resources, rather than from the textual content of the resource under analysis.

Finally, when asked if they found suggested tags with an unknown meaning, almost half of the participants (7 out of 15) answered positively, suggesting the need for an explanation tool, providing information about the suggested tags.

Overall, the feedback we received on the improvement with respect to a no-support situation was positive, and the support mechanisms we devised were perceived as valuable by the participants. In particular, the overall tagging experience was perceived as more engaging and more interesting, the availability of “private” tags was considered useful in order to reach a satisfactory final annotation, and system suggestions were (moderately) used. However, the evaluation results also show that improvements are required in order for such mechanisms to be more effective. They include the following.

- (i) The Like/Unlike feature should be empowered. In particular (i) other participants’ “likes” should be made immediately visible; (ii) information on how much a tag is “liked” should also be conveyed by the user interface.
- (ii) The management of “private” tags should be slightly revised taking into account the following points: (i) the possibility of saving the tags one “likes” in case they get removed by someone else should be kept; (ii) “private” tags should be available but visualized in a different way, less prominent with respect to shared ones.
- (iii) The suggestion functionality should be empowered in two directions: (i) the system should provide suggestions more closely related to the resource topic and also tags related to the goal of the collaboration (e.g., the very same resource could be differently tagged if used in a workspace devoted to the organization of a holiday in Ireland or to writing a scientific paper about Irish geology); this improvement could be obtained by taking into account the workspace context, mainly represented by the activity the workspace itself is devoted too and by the specific collaboration goals; the information about context and goals could be derived either from previously tagged resources or from some general workspace knowledge provided by users themselves as they develop their joint work; (ii) the system should be endowed with an explanation mechanism, providing users with information about the suggested tags; a preliminary work in this direction can be found in [46].

6. Conclusions

In this paper we described the design and a prototype implementation of a set of integrated mechanisms aimed at

supporting users in the collaborative semantic annotation of shared resources. The proposed approach includes the availability of *personal views*—that is, the possibility of keeping personal tags along with shared ones—and the suggestion of tags, based on both a syntactic and semantic analysis of the resource to be annotated.

We also presented a user evaluation, aimed at assessing such support mechanisms, in particular with respect to the problems highlighted by a previous user study, namely, the burden caused by the annotation activity and the possible disagreement among users about annotations. The evaluation showed that the devised support mechanisms actually improve the user experience and reduce both overload and possible disagreement. However, the evaluation results also provided us with interesting directions for our future work. In particular, besides revising the user interface according to the evaluation results, we plan to investigate the most effective way to enhance system suggestions by taking into account the workspace *context*, that is, a machine-readable representation of the activities the shared workspace is devoted to, and in particular of the *goals* they are aimed at.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] Indratmo and J. Vassileva, “A review of organizational structures of personal information management,” *Journal of Digital Information*, vol. 9, no. 26, pp. 1–19, 2008.
- [2] W. Jones, *Keeping Found Things Found*, Elsevier, Amsterdam, The Netherlands, 2007.
- [3] P. Ravasio and V. Tschertter, “Users’ theories of the desktop metaphor, or why we should seek metaphor-free interfaces,” in *Beyond the Desktop Metaphor*, V. Kaptelinin and M. Czerwinski, Eds., pp. 265–294, MIT Press, Cambridge, Mass, USA, 2007.
- [4] R. Boardman, R. Spence, and M. A. Sasse, “Too many hierarchies? The daily struggle for control of the workspace,” in *Human-Computer Interaction: Theory and Practice*, J. A. Jacko and C. Stephanidis, Eds., pp. 616–620, Erlbaum Associates, Mahwah, NJ, USA, 2003.
- [5] A. Goy, D. Magro, G. Petrone, and M. Segnan, “Semantic representation of information objects for digital resources management,” *Intelligenza Artificiale*, vol. 8, no. 2, pp. 145–161, 2014.
- [6] A. Goy, D. Magro, G. Petrone, C. Picardi, and M. Segnan, “Shared and personal views on collaborative semantic tables,” in *Semantic Web Collaborative Spaces*, P. Molli, J. Breslin, and M. E. Vidal, Eds., vol. 9507 of *Lecture Notes in Computer Science*, pp. 13–32, Springer, Heidelberg, Germany, 2016.
- [7] V. Kaptelinin and M. Czerwinski, Eds., *Beyond the Desktop Metaphor*, MIT Press, Cambridge, Mass, USA, 2007.
- [8] A. Adler, J. C. Nash, and S. Noël, “Evaluating and implementing a collaborative office document system,” *Interacting with Computers*, vol. 18, no. 4, pp. 665–682, 2006.
- [9] D. R. Karger, “Haystack: per-user information environments based on semistructured data,” in *Beyond the Desktop Metaphor*,

- V. Kaptelinin and M. Czerwinski, Eds., pp. 49–100, MIT Press, Cambridge, Mass, USA, 2007.
- [10] J. E. Bardram, “Activity-based computing: support for mobility and collaboration in ubiquitous computing,” *Personal and Ubiquitous Computing*, vol. 9, no. 5, pp. 312–322, 2005.
 - [11] S. Volda, E. D. Mynatt, and B. Macintyre, “Supporting activity in desktop and ubiquitous computing,” in *Beyond the Desktop Metaphor*, V. Kaptelinin and M. Czerwinski, Eds., pp. 195–222, MIT Press, Cambridge, Mass, USA, 2007.
 - [12] S. Houben, J. Vermeulen, K. Luyten, and K. Coninx, “Co-activity manager: integrating activity-based collaboration into the desktop interface,” in *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI '12)*, pp. 398–401, ACM, Capri Island, Italy, May 2012.
 - [13] P.-L. P. Rau, S.-H. Chen, and Y.-T. Chin, “Developing web annotation tools for learners and instructors,” *Interacting with Computers*, vol. 16, no. 2, pp. 163–181, 2004.
 - [14] J. Pearson, G. Buchanan, H. Thimbleby, and M. Jones, “The digital reading desk: a lightweight approach to digital note-taking,” *Interacting with Computers*, vol. 24, no. 5, pp. 327–338, 2012.
 - [15] J.-C. Jan, C.-M. Chen, and P.-H. Huang, “Enhancement of digital reading performance by using a novel web-based collaborative reading annotation system with two quality annotation filtering mechanisms,” *International Journal of Human Computer Studies*, vol. 86, pp. 81–93, 2016.
 - [16] L. Sauermann, A. Bernardi, and A. Dengel, “Overview and outlook on the semantic desktop,” in *Proceedings of the Workshop on the Semantic Desktop at ISWC 2005*, vol. 175, CEUR-WS, 2005.
 - [17] L. Drăgan, R. Delbru, T. Groza, S. Handschuh, and S. Decker, “Linking semantic desktop data to the web of data,” in *The Semantic Web—ISWC 2011*, L. Aroyo, C. Welty, H. Alani et al., Eds., vol. 7032 of *Lecture Notes in Computer Science*, pp. 33–48, Springer, Heidelberg, Germany, 2011.
 - [18] R. Koning, P. Grosso, and C. de Laat, “Using ontologies for resource description in the CineGrid Exchange,” *Future Generation Computer Systems*, vol. 27, no. 7, pp. 960–965, 2011.
 - [19] M. Uflacker and A. Zeier, “A semantic network approach to analyzing virtual team interactions in the early stages of conceptual design,” *Future Generation Computer Systems*, vol. 27, no. 1, pp. 88–99, 2011.
 - [20] D. K. Barreau and B. Nardi, “Finding and reminding: file organization from the desktop,” *ACM SIGCHI Bulletin*, vol. 27, no. 3, pp. 39–43, 1995.
 - [21] J. Teevan, C. Alvarado, M. S. Ackerman, and D. R. Karger, “The perfect search engine is not enough: a study of orienteering behavior in directed search,” in *Proceedings of the Conference on Human Factors in Computing Systems—Proceedings (CHI '04)*, pp. 415–422, Vienna, Austria, April 2004.
 - [22] F. Abel, N. Henze, D. Krause, and M. Kriesell, “Semantic enhancement of social tagging systems,” in *Web 2.0 & Semantic Web*, V. Devedžić and D. Gašević, Eds., vol. 6 of *Annals of Information Systems*, pp. 25–54, Springer, Boston, Mass, USA, 2009.
 - [23] R. Pak, S. Pautz, and R. Iden, “Information organization and retrieval: an assessment of taxonomical and tagging systems,” *Cognitive Technology*, vol. 12, no. 1, pp. 31–44, 2007.
 - [24] P. Dourish, W. K. Edwards, A. LaMarca et al., “Extending document management systems with user-specific active properties,” *ACM Transactions on Information Systems*, vol. 18, no. 2, pp. 140–170, 2000.
 - [25] A. Civan, W. Jones, P. Klasnja, and H. Bruce, *Better to Organize Personal Information by Folders or by Tags? The Devil is in the Details*, vol. 45, American Society for Information Science and Technology, Silver Spring, Md, USA, 2009.
 - [26] H. L. Kim, S. Scerri, J. G. Breslin, S. Decker, and H. G. Kim, “The state of the art in tag ontologies: a semantic model for tagging and folksonomies,” in *Proceedings of the 8th Annual International Conference on Dublin Core and Metadata Applications (DC '08)*, pp. 128–137, Berlin, Germany, September 2008.
 - [27] S. Lohmann, P. Díaz, and I. Aedo, “MUTO: the modular unified tagging ontology,” in *Proceedings of the 7th International Conference on Semantic Systems (I-SEMANTICS '11)*, pp. 95–104, Graz, Austria, September 2011.
 - [28] S. Angeletou, “Semantic enrichment of folksonomy tagspaces,” in *The Semantic Web—ISWC 2008*, A. Sheth, S. Staab, M. Dean et al., Eds., vol. 5318 of *Lecture Notes in Computer Science*, pp. 889–894, Springer, Heidelberg, Germany, 2008.
 - [29] A. Passant and P. Laublet, “Meaning of a tag: a collaborative approach to bridge the gap between tagging and linked data,” in *Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW '08)*, C. Bizer, T. Heath, K. Idehen, and T. Berners-Lee, Eds., vol. 369, CEUR-WS, Beijing, China, April 2008.
 - [30] A. Marchetti, M. Tesconi, F. Ronzano, M. Rosella, and S. Minutoli, “SemKey: a semantic collaborative tagging system,” in *Proceedings of the Workshop on Tagging and Metadata for Social Information Organization at WWW2007*, pp. 8–12, 2007.
 - [31] S. A. Golder and B. A. Huberman, “Usage patterns of collaborative tagging systems,” *Journal of Information Science*, vol. 32, no. 2, pp. 198–208, 2006.
 - [32] H. Cunningham, D. Maynard, K. Bontcheva et al., *Text Processing with GATE (Version 6)*, 2011, <https://gate.ac.uk/>.
 - [33] M. Erdmann, A. Maedche, H. P. Schnurr, and S. Staab, “From manual to semi-automatic semantic annotation: about ontology-based text annotation tools,” in *Proceedings of the Workshop on Semantic Annotation and Intelligent Content (COLING '00)*, pp. 79–85, ACL Press, Stroudsburg, Pa, USA, 2000.
 - [34] P. Fragkou, G. Petasis, A. Theodorakos, V. Karkaletsis, and C. Spyropoulos, “Boemie ontology-based text annotation tool,” in *Proceedings of the International Conference on Language Resources and Evaluation (LREC '08)*, N. Calzolari, K. Choukri, B. Maegaard et al., Eds., pp. 1273–1279, ELRA, 2008.
 - [35] K. Bontcheva, H. Cunningham, I. Roberts et al., “GATE Teamware: a web-based, collaborative text annotation framework,” *Language Resources & Evaluation*, vol. 47, no. 4, pp. 1007–1029, 2013.
 - [36] G. Petasis, “The SYNC3 collaborative annotation tool,” in *Proceedings of the International Conference on Language Resources and Evaluation (LREC '12)*, pp. 363–370, ELRA, 2012.
 - [37] V. Uren, P. Cimiano, J. Iria et al., “Semantic annotation for knowledge management: requirements and a survey of the state of the art,” *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 4, no. 1, pp. 14–28, 2006.
 - [38] O. Corcho, “Ontology based document annotation: trends and open research problems,” *International Journal of Metadata, Semantics and Ontologies*, vol. 1, no. 1, pp. 47–57, 2006.
 - [39] S. Borgo and C. Masolo, “Foundational choices in DOLCE,” in *Handbook on Ontologies*, S. Staab and R. Studer, Eds., pp. 361–381, Springer, Heidelberg, Germany, 2nd edition, 2009.
 - [40] A. Gangemi, S. Borgo, C. Catenacci, and J. Lehmann, *Task Taxonomies for Knowledge Content*, Metokis Deliverable D07, 2005.

- [41] D. Magro and A. Goy, "A core reference ontology for the customer relationship domain," *Applied Ontology*, vol. 7, no. 1, pp. 1–48, 2012.
- [42] A. Goy, G. Petrone, and M. Segnan, "A cloud-based environment for collaborative resources management," *International Journal of Cloud Applications and Computing*, vol. 4, no. 4, pp. 7–31, 2014.
- [43] A. Goy, D. Magro, G. Petrone, C. Picardi, and M. Segnan, "Ontology-driven collaborative annotation in shared workspaces," *Future Generation Computer Systems*, vol. 54, pp. 435–449, 2016.
- [44] G. Guest, A. Bunce, and L. Johnson, "How many interviews are enough? An experiment with data saturation and variability," *Field Methods*, vol. 18, no. 1, p. 24, 2006.
- [45] M. Crouch and H. McKenzie, "The logic of small samples in interview-based qualitative research," *Social Science Information*, vol. 45, no. 4, pp. 483–499, 2006.
- [46] A. Goy, D. Magro, G. Petrone, M. Rovera, and M. Segnan, "A semantic framework to enrich collaborative tables with domain knowledge," in *Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management Vol. 3: KMIS*, pp. 371–381, SciTePress, Setúbal, Portugal, 2015.

